

The clustering properties of the Lyman- α clouds

S. Cristiani^{1,3}, S. D’Odorico¹, V. D’Odorico^{2,3}, A. Fontana⁴, E. Giallongo⁴,
S. Savaglio¹

¹*European Southern Observatory, K. Schwarzschild Strasse 2, D-85748 Garching, Germany*

²*International School for Advanced Studies, SISSA, via Beirut 2-4, I-34013 Trieste, Italy*

³*Dipartimento di Astronomia, Università di Padova, vicolo dell’Osservatorio 5, I-35122 Padova, Italy*

⁴*Osservatorio Astronomico di Roma, via dell’Osservatorio, I-00040, Monteporzio, Italy*

Accepted ... Received ...; in original form 1995 December 28

ABSTRACT

We analyze the clustering properties of a high-resolution (~ 10 km/s) sample of about 1600 Lyman- α lines in the spectra of 15 quasars, obtained in the framework of an ESO key-programme with the addition of literature data. The two-point correlation function in the velocity space shows a significant signal on small velocity scales ($\Delta v \lesssim 300$ km/s) with amplitude and significance increasing with increasing column density. The correlation scale at $z \sim 3$ turns out to be $200 - 300 h_{50}^{-1}$ kpc. A trend of increasing correlation with decreasing redshift is apparent. The existence of over- and under-densities on scales of a few tens of Megaparsec is confirmed with a high confidence level and a number of possible structures are identified. The present observations are found to be consistent with models of gravitationally induced correlations. A continuity scenario between Lyman- α and metal systems emerges, with a suggested physical association between the Lyman- α clouds with $\log N_{HI} \gtrsim 14$ and the halos of protogalactic systems.

Key words: intergalactic medium – quasar absorption lines

1 INTRODUCTION

The Lyman- α “forest” of absorption lines seen in the spectra of high-redshift quasars is generally ascribed to pre-galactic or proto-galactic material. The study of the spatial distribution of these clouds offers a powerful tool to gain insight in the formation and evolution of structures up to redshifts 4-5, much larger than those reachable with observations of galaxies, limited at $z \lesssim 1$. The only other class of objects that at present can provide this type of information are the QSOs themselves. Lyman- α absorbers have, however, two definite advantages with respect to QSOs: they are much more numerous and, being less exotic objects, are expected to trace more “normal” matter.

The search for clustering of the Lyman- α lines has provided along the years results that have been defined disappointing or even controversial. Systematic studies of the distribution of redshifts in the QSO Lyman- α forest began in the early 1980’s with the work by Sargent et al. (1980), which concluded that no structures could be identified. Almost all the subsequent results have failed to detect any significant correlation on velocity scales $300 < \Delta v < 30000$ km s⁻¹ (Sargent et al. 1982; Bechtold 1987; Webb and Barcons 1991). On smaller scales ($\Delta v = 50 - 300$ km s⁻¹) there have been indications of weak clustering (Webb 1987; Rauch et

al. 1992; Chernomordik 1995; Cristiani et al. 1995), together with relevant non-detections (Pettini et al. 1990; Stengler and Webb 1993).

On the contrary, metal-line systems selected by means of the CIV doublet (Sargent et al. 1988) have been early recognized to show strong clustering up to 600 km s⁻¹, suggesting a different spatial distribution. In fact, the absence of power in the two-point correlation function has been claimed as a striking characteristics of the Lyman- α forest and has been used as a basic argument to develop a scenario of the Lyman- α absorbers as a totally distinct population with respect to metal systems and therefore galaxies.

2 THE DATABASE

The present analysis is based on data obtained in the framework of an ESO key-programme devoted to the study of QSO absorption systems at high redshifts. Up to now spectra of six QSOs, at resolutions between 9 and 14 km s⁻¹, have been reduced. The list of the objects, with emission redshifts ranging from 3.27 to 4.12 , is given in Table 1 together with the main characteristics of the sample of Lyman- α absorption lines. This constitutes a unique database, especially

Table 1. QSO Spectra from the ESO KP

QSO Name	z_{em}	FWHM (km s ⁻¹)	Mag	No of lines	Sample Limit log(N_{HI})
2126 – 15	3.27	11	V=17.3	106	13.3
2355 + 01	3.39	9	V=17.5	76	13.3
0055 – 26	3.67	14	V=17.5	187	13.3
1208 + 10	3.82	9	V=17.5	66	13.3
1108 – 07	3.95	9	R=18.1	38	13.3
0000 – 26	4.12	12	V=17.5	181	13.3

Other QSO Spectra from the literature

QSO Name	z_{em}	FWHM (km s ⁻¹)	Mag (V)	No of lines	Sample Limit log(N_{HI})
1331 + 17 ^a	2.10	18	16.9	55	13.0
1101 – 26 ^b	2.15	9	16.0	37	13.3
2206 – 19 ^c	2.56	6	17.3	47	13.3
1700 + 64 ^d	2.72	15	16.1	73	13.3
1946 + 76 ^e	3.02	10	15.8	72	13.3
0636 + 68 ^f	3.17	8	16.6	172	13.0
0302 – 00 ^f	3.29	8	17.6	151	13.0
0956 + 12 ^f	3.30	8	17.6	151	13.0
0014 + 81 ^f	3.41	8	16.5	164	13.0

References:

^a Kulkarni et al. 1995^b Carswell et al. 1991^c Rauch et al. 1993^d Rodríguez-Pascual et al. 1995^e Fan and Tytler 1994^f Hu et al. 1995

at high redshift, where the density of lines provides particular sensitivity to any clustering signal.

Spectra of the QSOs 2126 – 158, 0055 – 269, 0000 – 2619 have been already published (Giallongo et al. 1993; Cristiani et al. 1995; Savaglio et al. 1996). For 2126–158 new spectra have been added, significantly improving the S/N with respect to the published data. A complete description of the individual spectra will be given elsewhere: here we discuss the clustering properties of the Lyman- α sample.

All the spectra have been analyzed in a uniform way and all the lines have been fitted using our FITLYMAN program which is now available in the ESO-MIDAS package (Fontana and Ballester 1995). It performs a χ^2 minimization to derive the redshift z , the Doppler parameter b and the column density N for isolated lines and individual components of the blends. Whenever possible, the Lyman- β lines have been used to constrain the number of components in the strong saturated Lyman- α blends.

The S/N has been computed from the noise spectrum, and is typically greater than 10 per pixel element (corresponding to 15 per resolution element). In the regions near the QSO Lyman- α emissions and for the brighter quasars the S/N raises to $\sim 20 - 60$ per resolution element. Extensive simulations by various authors (Rauch et al. 1993; Fontana and Ballester 1995) show that no strong bias in the

derived statistical distributions is expected when the S/N is >10 .

To identify heavy element systems we have compared the lists of the observed lines with a catalog, derived from Morton (1991), containing the most frequently seen lines in QSO absorption spectra. As customary, we searched for significant excesses of identifications at all possible redshifts (Bahcall 1968).

We have complemented our data with other spectra available in the literature with similar resolution and redshift range, obtaining a final sample of 15 QSOs (in the following *extended sample*, see Table 1).

All the identified heavy element systems present in the Lyman- α forest have been removed from the extended sample, which consists of almost 1600 Lyman- α lines with observed column density $\log N_{HI} \geq 13$.

Lines affected by the proximity effect, i.e. within a proper distance of 10 Mpc from the emission redshift of each QSO, have been excluded from the clustering analysis.

We adopt throughout the value $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ for the Hubble parameter and $q_0 = 0.5$.

3 THE STATISTICAL ANALYSIS OF THE CLUSTERING

3.1 Small scales

As a first statistical tool to study the clustering properties of our sample of Lyman- α lines we have adopted the two-point correlation function (TPCF), defined as the excess, due to clustering, of the probability dP of finding a Lyman- α cloud in a volume dV at a distance r from another cloud:

$$dP = \Phi_{Ly\alpha}(z)dV[1 + \xi(r)] \quad (1)$$

where $\Phi(z)$ is the average space density of the clouds as a function of z . The TPCF is known to be a satisfactory estimator when used to investigate weak clustering on scales considerably smaller than the total interval covered by the data, which is the first domain we want to explore. The binning, intrinsic to this method, causes a loss of information, but the ease in visualizing its results and in including observational effects in the computing codes have made of the TPCF one of the favorite statistical estimators in cosmology.

In practice the observations provide the redshifts of the Lyman- α lines that, due to peculiar motions, are not immediately transformed in comoving distances. Therefore it is normally preferred to compute the TPCF in the velocity space, estimated as (Peebles 1980)

$$\xi(\Delta v) = \frac{N_{obs}(\Delta v)}{N_{exp}(\Delta v)} - 1 \quad (2)$$

where N_{obs} is the observed number of line pairs with velocity separations between Δv and $\Delta v + \epsilon_v$ and N_{exp} is the number of pairs expected in the same interval from a random distribution in redshift.

At the small velocity separations we are dealing with, the variation of the distance scale with cosmic time can be neglected and the velocity difference can be simply deduced from the redshift difference (Sargent et al. 1980)

$$\Delta v = \frac{c(z_2 - z_1)}{1 + (z_1 + z_2)/2} \quad (3)$$

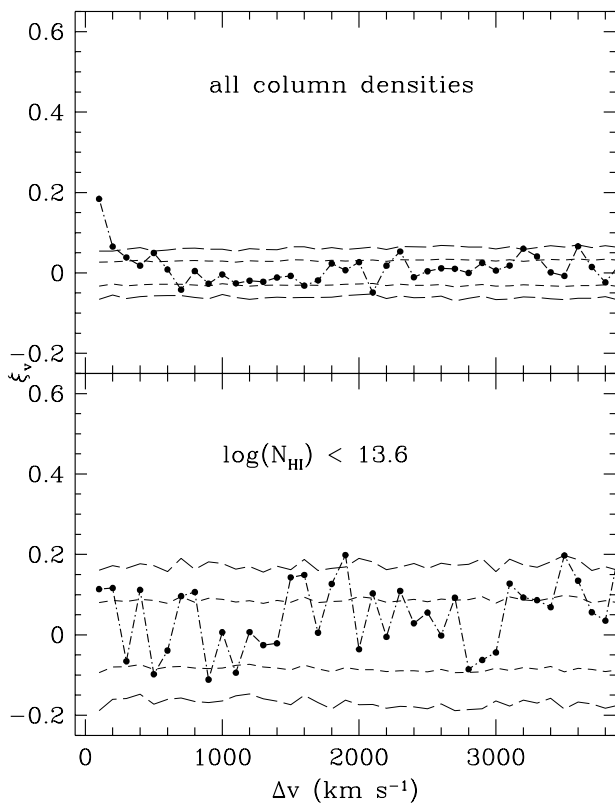


Figure 1. Two-point correlation function in the velocity space: *a*) (upper panel) for the complete sample of Lyman- α lines, *b*) (lower panel) for lines with column densities $< 10^{13.6} \text{ cm}^{-2}$. The short-dashed and long-dashed lines represent the 1σ and 2σ confidence limits for a Poissonian process

where Δv is the velocity of one cloud as measured by an observer in the rest-frame of the other.

In our line sample N_{exp} is obtained averaging 1000 numerical simulations of the observed number of redshifts, trying to account for all the relevant cosmological and observational effects. In particular the set of redshifts is randomly generated in the same redshift interval as the data according to the cosmological distribution $\propto (1+z)^\gamma$, where the best value of $\gamma = 2.65$ has been derived from a maximum likelihood analysis of the same database (Giallongo et al. 1996). The results are not sensitive to the value of γ adopted and even a flat distribution (i.e. $\gamma = 0$) gives values of ξ that differ typically by less than 0.02. Incomplete wavelength coverage due to gaps in the spectrum or line blanketing of weak lines due to strong complexes is also accounted for. Lines with too small velocity splittings, compared with the finite resolution or the intrinsic blending due to the typical line widths - the so called “line-blanketing” effect (Giallongo et al. 1996) -, are excluded in the estimate of N_{exp} .

The resulting correlation function for the full *extended sample* of Lyman- α lines is shown in the upper panel of Fig. 1. A weak but significant signal is present with $\xi \simeq 0.2$ in the 100 km s^{-1} bin: 739 pairs are observed while only 624 are expected for a random distribution, a 4.6σ deviation from poissonianity.

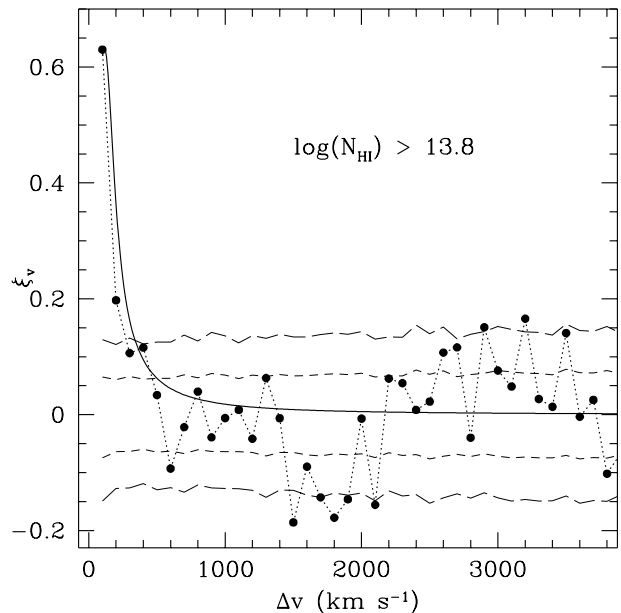


Figure 2. Two-point correlation function in the velocity space for lines with column densities $> 10^{13.8} \text{ cm}^{-2}$. Confidence limits as in Fig. 1. The continuous line shows the model described in Section 4, eq. 4, with $\gamma = 1.77$, $\sigma = 50 \text{ km/s}$, $r_{cl} = 180h_{50}^{-1} \text{ kpc}$ and $r_0 = 250h_{50}^{-1} \text{ kpc}$ at $z = 3$.

As in Cristiani et al. (1995), we have explored the variations of the clustering as a function of the column density. In Fig. 1b the correlation function for lines with $\log N_{HI} \leq 13.6$ is shown. All the evidence for clustering has disappeared. On the contrary, for lines with $\log N_{HI} \gtrsim 13.8$ (Fig. 2), the correlation function at $\Delta v = 100 \text{ km s}^{-1}$ shows a remarkable increase in amplitude ($\xi \simeq 0.6$) and significance: 234 pairs are observed while only 145 are expected for a random distribution, a more than 7σ deviation from poissonianity. No relevant feature other than the peak at small velocity separations is observed. In particular the evidence for anti-clustering on scales $\sim 600 - 1000 \text{ km s}^{-1}$, previously suggested by Meiksin and Bouchet (1995) on the basis of data of 0055 – 259 and 0014 + 813, is not confirmed. It should be noted that the assessment of the significance of a deviation from poissonianity is conceptually different when a velocity bin *given a priori* (as in the analysis of small-scale clustering) is analyzed, with respect to the case in which a deviation in *any* of the bins is searched for. While in the former situation the confidence levels shown in Figs. 1,2,4 (obtained by comparing the number of pairs observed in each bin with the distributions expected from simulations) are perfectly valid, in the latter case the significance of a positive or negative deviation in *any* of the bins is considerably less than what a naive comparison would at first glance suggest.

Fig. 3 shows the variation of the amplitude of the two-point correlation as a function of the column density threshold for the sample of the Lyman- α lines. On the same plot are reported for comparison the TPCF’s derived for high and low column density CIV metal systems from the works by Petitjean and Bergeron (1994) and by Songaila and Cowie (1996), respectively. A reference value of $\log(N_{HI}) \sim 17 \pm 0.5$ has been assigned to high column density CIV metal systems

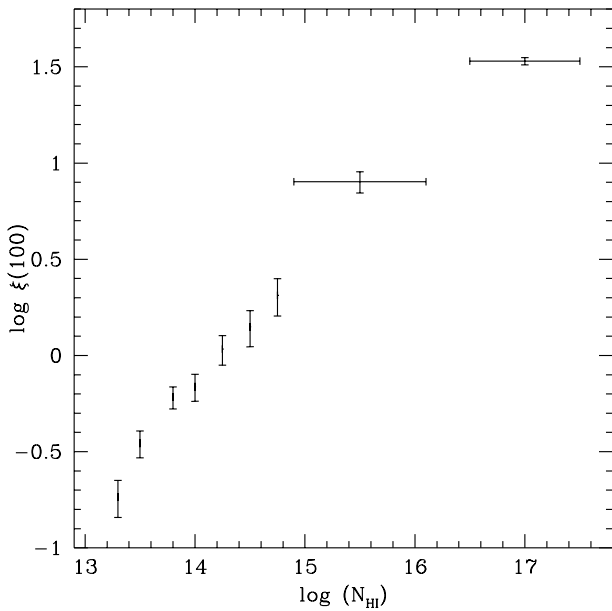


Figure 3. Evolution of the amplitude of the two-point correlation function as a function of the column density threshold for the sample of the Lyman- α lines. The two points in the upper-right side of the picture show the correlation of the CIV metal systems (see text)

on the basis of observations of 0420-388 (Atwood et al. 1985; Petitjean and Bergeron 1994) and of 2126-158 (Giallongo et al. 1993; Petitjean and Bergeron 1994). For the low column density CIV systems a value of $\log(N_{HI}) \sim 15.5 \pm 0.6$ is derived from Songaila and Cowie (1996).

An extrapolation of the increasing amplitude trend observed for the TPCF of the Lyman- α lines would easily intercept the corresponding estimates derived from the CIV metal systems. We stress also the similarity between the shapes of the TPCF's of Lyman- α and CIV systems, when observed at comparable resolution (cfr. Fig. 4b of Petitjean and Bergeron 1994, Womble et al. 1995, Fig. 6 of Songaila and Cowie 1996) and the fact that a trend of increasing correlation with increasing column density is observed also in CIV systems, for which the strong systems appear usually in clumps of many components, while the weak doublets (generally optically thin at the Lyman limit) are typically made of an isolated, single component (Bergeron 1995).

In a recent paper, Fernandez-Soto et al. (1996) have investigated the clustering properties of the Lyman- α clouds on the basis of the corresponding CIV absorptions, suggesting that CIV may resolve better the small-scale velocity structure that cannot be fully traced by Lyman- α lines. As a consequence, the estimates of the TPCF of the Lyman- α absorbers, although the effects of the non-negligible width of the lines (the “line-blanketing”) are considered in our simulations, should be regarded as a lower limit to the real clustering amplitude. However, it is not straightforward to translate the properties of the CIV absorbers in the corresponding ones of the Lyman- α absorbers, since observations (e.g. Prochaska and Wolfe 1996) show that the velocity structures of high and low-ionization species are often different. Since the underestimation of the Lyman- α TPCF

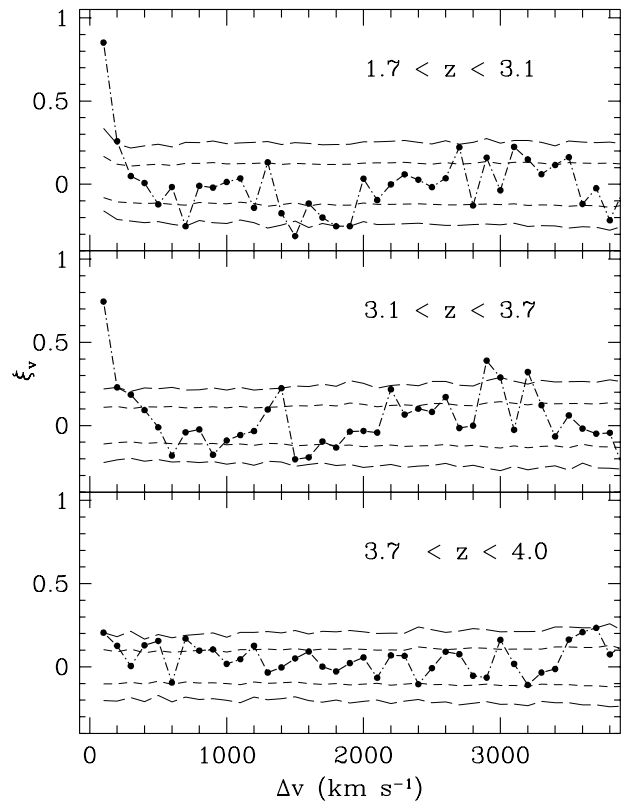


Figure 4. Evolution of the two-point correlation function with redshift for the sample of the Lyman- α lines with column densities above $\log(N_{HI}) = 13.8$. The three redshift intervals $1.7 < z < 3.1$, $3.1 < z < 3.7$, $3.7 < z < 4.0$ (upper, middle and lower panel, respectively) have been chosen in a way that the confidence limits, estimated as in Fig. 2-3, turn out to be roughly the same in all three ranges

would be more severe at larger column densities, the trend of an increasing correlation with increasing column density appears to be real, from the lowest column densities up to those corresponding to the strongest metal systems.

In Fig. 4 we study the evolution of the TPCF with the redshift for the sub-sample of Lyman- α lines with column densities $\log(N_{HI}) > 13.8$. The amplitude of the correlation at 100 km s^{-1} decreases with increasing redshift from 0.85 ± 0.14 at $1.7 < z < 3.1$, to 0.74 ± 0.14 at $3.1 < z < 3.7$ and 0.21 ± 0.14 at $3.7 < z < 4.0$. Unfortunately, HST data are still at too low-resolution (e.g. Bahcall et al. 1995) or are too scanty (Brandt et al. 1993) to allow a meaningful comparison with the present data. Nonetheless, the existence at low-redshifts of clumps of Lyman- α absorption lines clustered around metal-line systems, reported by Bahcall et al. (1995) as a resemblance with the properties of galaxies and of metal absorption systems, is suggestive that the trend illustrated in Fig. 4 persists at lower redshifts, as confirmed by the measure of the Lyman- α TPCF, $\xi = 1.8^{+1.6}_{-1.2}$, on scales of $200 - 500 \text{ km/s}$ at $0 < z < 1.3$ (Ulmer 1996).

Table 2. QSO Spectra used in the search for voids

QSO Name	z_{em}	minimum void size Δr (Mpc)	Number of detected voids
0636 + 68	3.17	15.9	1
2126 – 15	3.27	24.7	0
0302 – 00	3.29	17.5	0
0956 + 12	3.30	16.9	0
2355 + 01	3.39	13.1	0
0014 + 81	3.41	15.8	0
0055 – 26	3.67	14.9	2
1208 + 10	3.82	8.5	0
0000 – 26	4.12	6.9	1

3.2 Large scales

The TPCF is known to be inefficient for detecting structures on scales comparable with respect to the dimension of the sample (e.g. Pando and Fang 1996). For this reason to investigate the clustering properties of the Lyman- α absorbers on scales of several comoving megaparsecs, searching for over- and under-densities of lines significantly deviating from a Poissonian distribution, we have resorted to different statistical techniques. The aim was not only to assess the existence of non-random fluctuations on large scales, but also to identify the regions of the spectra giving origin to such deviations.

3.2.1 Voids

A typical way of looking for non-random fluctuations of the line density on large scales is the search for voids. Voids in the Lyman- α forest provide a test for models of the large-scale structure in the Universe and of the homogeneity of the UV ionizing flux. Previous searches for megaparsec-sized voids have produced a few claims (Crotts 1989; Dobrzycki and Bechtold 1991b), but uncertainties in the line statistics strongly influence the probability estimate (Ostriker 1988). High resolution data, less affected by blending effects, are ideal also for the study of voids, but great care has to be taken in the statistical approach to avoid the pitfalls of “a posteriori statistics”. Detections of voids in individual cases are interesting but a more general approach, assessing more quantitatively how common is the phenomenon, is to be preferred. Differences in the redshifts and S/N (and consequently in the line density) make difficult a global and uniform usage of our extended sample in this sense. Therefore we have designed the following procedure: for each of the objects listed in Table 1 and for lines with $\log(N_{HI}) \geq 13.3$ we have estimated through Montecarlo simulations the void size Δr (in comoving coordinates) for which the probability to find at least one void $\geq \Delta r$ is 0.05. The objects for which the density of lines and the absence of spectral gaps allow a meaningful analysis are listed in Table 2. Then we have searched each spectrum for voids of dimensions equal or larger than Δr . In 3 out of 9 cases at least one void was detected (two voids for 0055–26). The binomial probability corresponding to such occurrences is $8 \cdot 10^{-3}$.

The four voids reported in Table 2 have been found in

the spectra of 0636+68 at 4339.5–4361.6 Å (16.1 Mpc size), 0055 – 26 at 4979.05 – 5016.6 and 5188.4 – 5223.8 Å (22.3 and 19.8 Mpc, respectively) and 0000 – 26 at 5627 – 5644 Å (8.5 Mpc). The well known void at 5050 – 5100 Å in the spectrum of 0302 – 00 (Dobrzycki and Bechtold 1991b) has not been found because the spectrum of Hu et al. (1995) does not cover this region of the Lyman forest. From the published spectra and from Fig. 5 it is apparent that the regions corresponding to the voids are not completely devoid of lines: weak absorptions are observed within the voids. This agrees with low-redshift observations (Shull, Stocke and Penton 1996) that have shown that in the local Universe voids are not entirely devoid of matter.

Even if underdense regions are statistically significant in our sample, the filling factor is rather low: the voids reported in Table 2 cover only about 2% of the available line-of-sight path-length, confirming what already noticed by Carswell and Rees (1987).

3.2.2 Over- and under-densities of lines

Voids are just an extreme case of spectral regions showing an under-density of lines. Various theoretical reasons prompt to tackle the issue of over- and under-densities of lines from a more general point of view: the typical signature of a “proximity effect” due to a foreground quasar is a lack of weak lines in the Lyman forest, rather than lack of lines in general (Dobrzycki and Bechtold 1991a); the relative filling factor of under-densities and over-densities may provide a constraint on the theories of structure formation.

We have analysed the spectra of the quasars reported in Tab. 1 with a counts-in-cells technique searching for over- and under-densities of lines with $\log(N_{HI}) \geq 13.3$ on scales from 10 to 80 Mpc and comparing the observed counts with Montecarlo simulations in order to assess the significance of the deviations. On smaller scales the shot noise is too large, on larger scales the “integral constraint”, forcing the simulated number of lines to be equal to the observed one, prevents from the possibility of detecting any deviation. The threshold to define significant a deviation (for example in excess) in a given spectral interval has been set in a way that at a given scale for a given quasar there is a 0.05 probability of observing at least one deviation of this type on the whole spectrum for a locally Poissonian distribution. In Table 3 we report the results of the analysis. An asterisk denotes that a meaningful search at that particular scale has not been possible for the given object.

5 QSOs out of 15 show at least one over-density in their spectrum at 10 Mpc scales, corresponding to a binomial probability of $6 \cdot 10^{-4}$ of being drawn from a poissonian distribution of lines. 4 QSOs show at least one over-density of lines at 20 and 30 Mpc scales, corresponding to a binomial probability of $5 \cdot 10^{-3}$. At larger scales the number of significant over-densities decreases. None is observed at 80 Mpc. Over-densities detected at one scale tend to persist at larger scales, analogously to what observed in the wavelet identification of structures from galaxy counts (Slezak, Bijaoui and Mars 1990).

Under-densities appear to be roughly as common as over-densities, once the lack of sensitivity at lower redshifts and smaller scales, due to the low density of lines, is taken into account. In practice in Table 4 the voids of Table 2

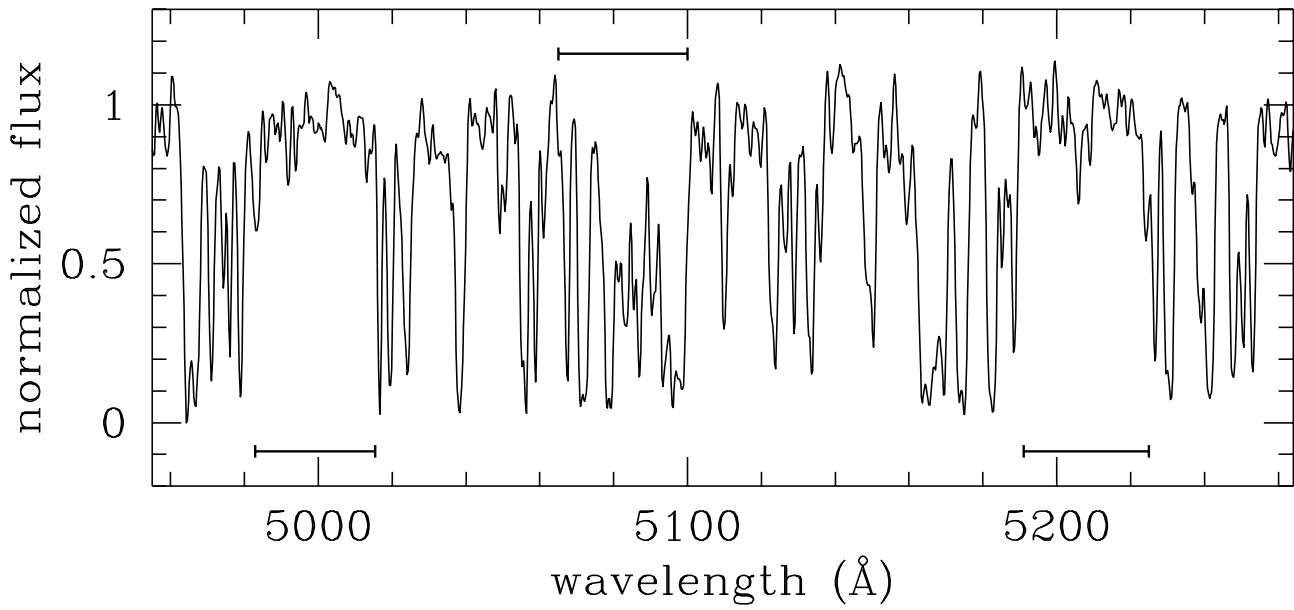


Figure 5. Over and under-densities of lines at a scale of 20 Mpc in the spectrum of Q0055 – 26. The positions of two voids, at 5000 and 5200 Å respectively (see Tab. 2), are shown together with and over-density at 5080 Å. The relatively low resolution of the plot does not allow to fully visualize the significance of the over-density: 15 lines with $\log(N_{HI}) \geq 13.3$ are observed while only 7 are expected.

Table 3. Over-densities of lines found in QSO spectra

Scale	10 Mpc	20 Mpc	30 Mpc	40 Mpc	80 Mpc
QSO Name	over-density center (Å)				
1331 + 17	3494	3493	3490	3490	-
	3605	3609			
1101 – 26	3742	-	3743	-	-
2206 – 19	-	-	-	-	-
1700 + 64	-	-	-	-	-
1946 + 76	4427	4435	-	-	-
0636 + 68	-	-	-	-	-
2126 – 15	-	-	-	-	-
0302 – 00	-	-	4696	-	-
0956 + 12	-	-	-	-	-
2355 + 01	-	5217	-	-	-
0014 + 81	-	-	-	-	-
0055 – 26	5085	5080	-	-	-
1208 + 10	5621	-	5645	-	*
1108 – 07	-	-	-	5839	*
0000 – 26	-	-	-	-	-

are recovered, except for the void of 0636 + 68, the least significant case.

4 DISCUSSION

As shown by Fernandez-Soto et al. (1996), a TPCF approach applied - as in the present work - directly to the Lyman- α lines tends to underestimate their clustering properties. However, the very fact that a signal is found, in spite of all the limitations described by Fernandez-Soto et al. (1996), is a significant result and, even if regarded as lower limits, the

Table 4. Under-densities of lines found in QSO spectra

Scale	10 Mpc	20 Mpc	30 Mpc	40 Mpc	80 Mpc
QSO Name	under-density center (Å)				
1331 + 17	*	*	3707	-	-
1101 – 26	*	*	3679	-	3670
2206 – 19	*	4238	-	-	-
1700 + 64	*	-	-	-	-
1946 + 76	*	-	-	4799	-
0636 + 68	*	-	-	-	-
2126 – 15	*	-	-	-	-
0302 – 00	*	-	-	-	-
0956 + 12	*	-	-	-	-
2355 + 01	*	-	-	-	-
0014 + 81	*	-	-	-	-
0055 – 26	5263	5001	5007	4988	-
		5208	5406		
1208 + 10	-	-	-	-	-
1108 – 07	-	-	-	-	*
0000 – 26	5637	-	-	-	-

measured clustering amplitudes allow to draw interesting inferences.

In this way, the observed clustering properties are qualitatively consistent with a scenario of gravitationally induced correlations. The dependence of the clustering amplitude on the column density, reported in subsection 3.1, is easily explained by models involving biasing for the formation of structures in the universe. Objects associated with the stronger potential wells are expected to be more clustered and the HI column density is naturally related to the depth of the well (and to the associated mass).

The trend of increasing correlation with decreasing red-

shift (Fig. 4) is also a strong prediction of any model of structure formation based on gravitational instability. On the contrary, theories of explosive structure formation (Vishniac, Ostriker and Bertschinger 1985; Weinberg, Ostriker and Dekel 1989) expect a velocity correlation either unchanging or diminishing with time.

The existence of a roughly equal number of over and under-densities on scales 10 – 40 Mpc is easily understandable in terms of the linear theory of the evolution of the perturbations, that is a plausible approximation at such relatively high redshifts: gravity has not yet had time to give a significant skewness to the (under)over-density distribution. Besides, almost any hierarchical clustering scenario would expect that at $z \sim 2 - 4$ gravity has not yet had enough time to transfer power on 80 Mpc scales and give origin to significant over or under-densities.

Numerical simulations in the framework of a CDM model, including photo-ionization and cooling of the baryonic component, have reproduced to a remarkable extent the observed correlation function and also its dependence on the column density of the lines (Mücket et al. 1996).

In this way the observed clustering properties (in particular the correlation between clustering amplitude and column density) become part of a new continuity scenario that rests on two other pieces of evidence: the metallicities of the order 10^{-2} observed for Lyman- α clouds with $\log N_{HI} > 14$ (Cowie et al. 1995; Tytler et al. 1995) and the volume density and cosmological evolution of the same clouds found to be similar to those of the damped systems (Giallongo et al. 1996). All together this suggests a physical association between the Lyman- α clouds with $\log N_{HI} > 14$ and the halos of protogalactic systems. At low redshift a considerable fraction of the Lyman- α lines has been observed to be associated with luminous galaxies and the local large-scale structure (Lanzetta, Bowen, Tytler et al. 1995; Le Brun and Bergeron 1996).

The typical column density below which no clustering of the Lyman- α lines is observed ($\log(N_{HI}) \sim 13.6$, Fig. 1) corresponds to the position of the break in their column density distribution (Giallongo et al. 1996; Hu et al. 1995), which has been identified with the transition from a variety of systems in various stages of gravitational infall and collapse (or even under-densities) to gas associated with star forming galaxies (Zhang, Anninos and Norman 1996; Mücket et al. 1996).

Following Heisler, Hogan and White (1989) and Cristiani et al. (1995), it is possible to derive the three-dimensional spatial autocorrelation function of the clouds ξ_r , from the velocity correlation ξ_v , assuming a Gaussian distribution of the peculiar motions of the clouds with respect to the Hubble flow and a power-law form for the spatial correlation function, as observed for galaxies.

$$\xi_v = \int_0^\infty Hdr \xi(r) P(v|r) \propto \int_{r_{cl}}^\infty \frac{Hdr}{\sigma} \left(\frac{r}{r_o}\right)^{-\gamma} \times \left\{ \exp\left[-\frac{(Hr-v)^2}{2\sigma^2}\right] + \exp\left[-\frac{(Hr+v)^2}{2\sigma^2}\right] \right\} \quad (4)$$

At small velocity splittings, the correlation scale r_o depends mainly on the cloud sizes r_{cl} and on the velocity dispersion σ assumed. The very fact that a significant correlation in velocity is observed results in an upper limit on

the cloud proper sizes of the order of $200h_{50}^{-1}$ kpc. It is interesting to compare this upper limit with the lower limits on the cloud dimensions inferred from the observations of QSO pairs or gravitational lenses (Smette et al. 1992; Smette et al. 1995; Bechtold et al. 1994; Dinshaw et al. 1995), $r_{cl} \gtrsim 100h_{50}^{-1}$ kpc. Indeed, the trend observed in Fig. 4 is consistent, in the framework of a simple hierarchical clustering model predicting that r_o scales with the redshift according to the relation $r_o(z) = r_o(0)(1+z)^{-5/3}$, with large cloud sizes ($r_{cl} \lesssim 200h_{50}^{-1}$ kpc) and low velocity dispersions (few tens km/s).

Assuming an index $\gamma = 1.77$ for the power-law spatial correlation function as commonly measured for galaxies, relatively low velocity dispersions (< 100 km/s) as suggested by observations of quasar pairs (e.g. Dinshaw et al. 1994) and cloud sizes of $200h_{50}^{-1}$ kpc, implies, to reproduce the observed correlation in the velocity space, a correlation scale at $z \sim 3$ of $200 - 300h_{50}^{-1}$ comoving kpc. This amplitude extrapolates, according to the $(1+z)^{-5/3}$ scaling, to $2 - 3h_{50}^{-1}$ Mpc at the present epoch, to be compared with the values 10 and $7.5h_{50}^{-1}$ Mpc observed for present day optical (de Lapparent, Geller and Huchra 1988) and IRAS (Fisher, Davis, Strauss et al. 1994) galaxies, respectively.

A spatial correlation function with a correlation scale $r_o(z=0) \sim 3$ Mpc, convolved with a velocity dispersion $\sigma \sim 50$ km/s, would result in a correlation function in the velocity space non dissimilar with respect to what is observed for CIV systems (Womble et al. 1995; Songaila and Cowie 1996) if a cloud radius $\sim 15h_{50}^{-1}$ proper kpc is assumed. Such a value appears considerably smaller than what is commonly inferred for galactic halos giving origin to metal-rich absorption systems ($r \sim 75h_{50}^{-1}$ kpc, Bergeron 1995), but certainly at least part of the small-scale correlation observed for CIV systems is due to the motion of clouds of size $\ll 75$ kpc within individual galaxy halos. The observed correlations in the velocity space might then arise from the same spatial correlation function both for Lyman- α and metal systems, with different degrees of amplitude suppression due to the different sizes of the regions giving origin to the Lyman- α and metal absorption, a well-known phenomenon occurring when the cloud sizes are non-negligible with respect to the correlation scale (Bajtlik 1995).

5 CONCLUSIONS

The analysis of a large database (~ 1600 lines) of Lyman- α absorptions observed at ~ 30000 resolution in the spectra of 15 quasars has produced the following results:

- (i) clustering on small velocity scales ($\Delta v \lesssim 300$ km/s) has been measured at a 5σ confidence level for the full line sample, which has typically $\log(N_{HI}) \gtrsim 13$, with $\xi_v(100 \text{ km/s}) = 0.2$
- (ii) the clustering amplitude increases with increasing column density of the lines: $\xi_v(100 \text{ km/s}, \log(N_{HI}) \geq 13.8) = 0.6$. No significant clustering is observed for lines with $\log(N_{HI}) \lesssim 13.6$. An extrapolation of this trend to column densities typical of metal systems is consistent with the clustering observed for CIV lines
- (iii) a trend of decreasing clustering with increasing redshift is observed

(iv) a number of over- and under-densities of lines (including 4 voids) have been identified on scales of a few tens of Mpc, confirming at a high confidence level the reality of deviations from a poissonian distribution of lines on such relatively large scales. The filling factor of such structures is however low: a few percent of the total line-of-sight path-length.

Although the TPCF approach adopted in the present paper tends to underestimate the small-scale clustering amplitudes, whose measured values have to be regarded as lower limits, the following conclusions can be drawn:

- (i) the observed clustering properties appear qualitatively consistent with a scenario of gravitationally induced correlations
- (ii) simple considerations about the two-point correlation function in real and velocity space allow to deduce an upper limit to the cloud sizes $r_{cl} \lesssim 200h_{50}^{-1}$ proper kpc and a correlation scale $r_o = 200 - 300h_{50}^{-1}$ comoving kpc at $z \sim 3$.
- (iii) a continuity scenario between Lyman- α and metal systems is suggested, with a physical association between the Lyman- α clouds with $\log N_{HI} \gtrsim 14$ and the halos of protogalactic systems.

ACKNOWLEDGMENTS

We thank J. Bergeron, S. Matarrese, L. Moscardini, R. Vio, S. White, for helpful discussions and P. Andreani for a critical reading of the original version of the paper. SC acknowledges the support of the ASI contract 94-RS-107.

REFERENCES

- Atwood B., Baldwin J.A., Carswell R.F., 1975, *Astrophys. J.*, **292**, 58
- Bahcall J., 1968, *Astrophys. J.*, **153**, 679
- Bahcall J., Bergeron J., Boksenberg A., et al., 1995, preprint
- Bajtlik S., 1995, in Meylan ed., Proc. Workshop QSO absorption lines. Springer, Berlin, p. 337
- Bechtold J., 1987, in J. Bergeron et al. eds., Proc. Third IAP Workshop, High Redshift and Primeval Galaxies. Editions Frontieres, Gif-sur-Yvette, p. 397
- Bechtold J., Crotts A. P. S., Duncan R. C., Fang Y., 1994, *Astrophys. J.*, **437**, L83
- Bergeron J., 1995, in Meylan ed., Proc. Workshop QSO absorption lines. Springer, Berlin, p. 127
- Brandt J. C., Heap S. R., Beaver E. A., et al., 1995, *Astron. J.*, **105**, 831
- Carswell R. F., Rees M. J., 1987, *Mon. Not. R. astr. Soc.*, **224**, 13p
- Chernomordik V.V., 1995, *Astrophys. J.*, **440**, 431
- Cowie L.L., Songaila A., Kim T., Hu E.M., 1995, *AJ*, **109**, 1522
- Cristiani S., D'Odorico S., Fontana A., Giallongo E., Savaglio S., 1995, *Mon. Not. R. astr. Soc.*, **273**, 1016
- Crotts A. P. S. 1989, *Astrophys. J.*, **336**, 550
- de Lapparent V., Geller M.J., Huchra J.P., 1988, *Astrophys. J.*, **332**, 44
- Dinshaw N., Foltz C.B., Impey C.D., Weymann R.J., Morris S.L. 1995, *Nature*, **373**, 223
- Dobrzycki, A., Bechtold, J. 1991a, in Crampton ed., Proc. Workshop The space distribution of quasars, ASP Conference series 21, p. 272
- Dobrzycki, A., Bechtold, J. 1991b, *Astrophys. J.*, **377**, L69
- Fan X., Tytler D., 1994, *Astrophys. J. Suppl.*, **94**, 17
- Fernandez-Soto A., Lanzetta K.M., Barcons X., Carswell R.F., Webb J.K., Yahil A., *Astrophys. J.*, **460**, L85
- Fisher K. B., Davis M., Strauss M.A., Yahil A., Huchra J., 1994, *Mon. Not. R. astr. Soc.*, **266**, 50
- Fontana A., Ballester P., 1995, *ESO Messenger*, **80**, 37
- Giallongo E., Cristiani S., Fontana A., Trevese D. 1993, *Astrophys. J.*, **416**, 137
- Giallongo E., Cristiani S., D'Odorico S., Fontana A., Savaglio S., 1995, *Astrophys. J.*, ESO preprint No. 1132
- Heisler, J., Hogan, C. J., White, S. D. M. 1989, *Astrophys. J.*, **347**, 52
- Hu E. M., Kim T., Cowie L. L., Songaila A., 1995, preprint
- Kulkarni V. P., Huang K., Green R. F., Bechtold J., Welty D. E., York D. G., 1995, preprint
- Lanzetta K. M., Bowen D. B., Tytler D., Webb J. K., 1995, *Astrophys. J.*, **442**, 538
- Le Brun V., Bergeron J., Boisse P., 1996, *Astron. Astrophys.*, **306**, 691
- Meiksin A., Bouchet F. R., 1995, *Astrophys. J.*, **448**, L85
- Morton D.C., 1991, *Astrophys. J. Suppl.*, **77**, 119
- Mückel J. P., Petitjean P., Kates R. E., Riediger R., 1996, *Astron. Astrophys.*, **308**, 17
- Ostriker J. P., Bajtlik S., Duncan R. C., 1988, *Astrophys. J.*, **327**, L35
- Pando J., Fang L.Z., 1996, *Astrophys. J.*, **459**, 1
- Peebles P.J.E., 1980, The Large Scale Structure of the Universe. Princeton Univ. Press, Princeton
- Petitjean P., Bergeron J., 1994, *Astron. Astrophys.*, **283**, 759
- Pettini M., Hunstead R. W., Smith L. J., Mar D. P. 1990, *Mon. Not. R. astr. Soc.*, **246**, 545
- Prochaska J.X., Wolfe A.M., 1996, babbage astro-ph/9604042
- Rauch M., Carswell R. F., Chaffee F. H., Foltz C. B., Webb J. K., Weymann R. J., Bechtold J., Green R. F. 1992, *Astrophys. J.*, **390**, 387
- Rauch M., Carswell R. F., Webb J. K., Weymann R.J., 1993, *Mon. Not. R. astr. Soc.*, **260**, 589
- Rodríguez-Pascual P.M., De La Fuente A., Sanz J.L., Recondo M.C., Clavel J., Santos-Lleó M., Wamsteker W., 1995, *Astrophys. J.*, **448**, 575
- Sargent W. L. W., Young P. J., Boksenberg A., Tytler D. 1980, *Astrophys. J. Suppl.*, **42**, 41
- Sargent W. L. W., Young P. J., Schneider D. P., 1982, *Astrophys. J.*, **256**, 374
- Sargent W. L. W., Boksenberg A., Steidel C. C. 1988, *Astrophys. J. Suppl.*, **68**, 539
- Savaglio S., Cristiani, D'Odorico S., Fontana A., Giallongo, E., Molaro P., 1996, babbage astro-ph/9606063
- Shull J.M., Stocke J.T., Penton S., 1996, *Astron. J.*, **111**, 72
- Slezak E., Bijaoui A., Mars G., 1990, *Astron. Astrophys. Suppl.*, **227**, 301
- Smette A., Surdej J., Shaver P.A. et al., 1992, *Astrophys. J.*, **389**, 39
- Smette A., Robertson J.G., Shaver P.A., Reimers D., Wisotzki L., Köhler Th., 1995, *Astron. Astrophys. Suppl.*, **113**, 199
- Songaila A., Cowie L. L., babbage astro-ph/9605102
- Stengler-Larrea E. A., Webb J. K. 1993, in Chincarini G., Iovino A., Maccacaro T., Maccagni D. eds., Observational Cosmology, ASP Conference Series, **51**, 591
- Tytler D., Fan X.-M., Burles S., Cottrell L., Davis C., Kirkman D., Zuo L., 1995, in *QSO Absorption Lines*, Proc. ESO Workshop, ed. G.Meylan (Heidelberg: Springer), p.289
- Ulmer A., 1996, *Astrophys. J.* submitted
- Vishniac E.T., Ostriker J.P., Bertschinger E., 1985, *Astrophys. J.*, **291**, 399
- Webb J. K., 1987, in Hewett A., Burbidge G., Fang L. Z. eds., Proc. IAU Symp. 124, Observational Cosmology. Reidel, Dordrecht, p. 803

- Webb J. K., Barcons X., 1991, *Mon. Not. R. astr. Soc.*, **250**, 270
- Weinberg D.M., Ostriker J.P., Dekel A., 1989, *Astrophys. J.*, **336**, 9
- Womble D.S., Sargent W.L.W., Lyons R.S., 1995, babbage astro-ph/9511035
- Zhang Y., Anninos P., Norman M.L., 1996, *Astrophys. J.*, **459**, 12